

Integrated Expression Analysis May Support Serine/Threonine Kinases as Common Hub Genes in Breast Cancer

Mohammad Soleiman Ekhtiyari¹, MSc;^{ORCID}
Mostafa Ghaderi-Zefrehei², PhD;^{ORCID} Zahra Mogharari³, MSc; Maryam Yousefi⁴, PhD;
Ali Bigdeli⁵, MSc; Effat Nasre Esfahani⁶, PhD;
Hamed Shahriarpour⁵, MSc; Bluma J. Leschm⁷, PhD

¹Division of Biochemistry, Faculty of Veterinary Medicine, University of Tabriz, Tabriz, Iran;

²Department of Animal Genetics, Faculty of Agriculture, Yasouj University, Yasouj, Iran;

³Department of Genetics, Faculty of Agriculture, Shahed University, Tehran, Iran;

⁴Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran;

⁵Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran;

⁶Department of Animal Sciences, Payame Noor University, Tehran, Iran;

⁷Department of Genetics, Department of Obstetrics, Gynecology and Reproductive Sciences, Yale Cancer Center, Yale School of Medicine, New Haven, CT, USA

Correspondence:

Mostafa Ghaderi-Zefrehei, PhD;
Pasdaran St., Central Building of Yasouj University, Postal Code: 75918-74934, Yasuj, Iran

Email: mghaderi@yu.ac.ir

Received: 12 November 2024

Revised: 11 December 2024

Accepted: 27 January 2025

What's Known

- Serine/threonine kinases, including *NEK2* (NimA-Related Protein Kinase 2), *MELK* (Maternal Embryonic Leucine Zipper Kinase), *PLK1* (Polo Like Kinase 1), *AURKB* (Aurora Kinase B), and *CHEK1* (Checkpoint Kinase 1), are linked to breast cancer progression.
- These kinases play a critical role in regulating the cell cycle and are associated with poor overall survival in breast cancer patients.

What's New

- This study identifies a cluster of five serine/threonine kinases as hub genes in breast cancer through integrated transcriptomic analysis.
- The structural similarities in their binding sites suggest potential for targeted therapies aimed at these conserved domains in breast cancer treatment

Abstract

Background: Breast cancer (BC) is the most common cancer affecting women worldwide. There is a strong need to identify molecular pathways that might represent effective therapeutic targets.

Methods: We conducted a large-scale transcriptomic analysis using publicly available datasets from the NCBI GEO and TCGA databases. Microarray datasets (GSE161533, GSE162228, GSE70947, and GSE139038) and RNA-Seq data were analyzed to identify differentially expressed genes (DEGs) using cut-off criteria of adjusted $P < 0.05$ and $|\log_2FC| > 1$. Gene co-expression networks were constructed using Weighted Gene co-expression Network Analysis (WGCNA) in R (version 1.68), followed by hub gene identification with STRING and MCODE tools. Functional enrichment was further explored through Gene Ontology analysis.

Results: Two regulatory modules enriched in cancer datasets were identified from both microarray and RNA-Seq analyses, corresponding to a network of 85 genes, compared to a distinct network of 474 genes enriched in control tissue samples. Further analyses to identify densely connected gene clusters within these networks revealed a cluster containing 29 cancer-related genes that included five hub gene candidates encoding serine/threonine kinase family proteins: NimA-Related Protein Kinase 2 (*NEK2*), Maternal Embryonic Leucine Zipper Kinase (*MELK*), Polo Like Kinase 1 (*PLK1*), Aurora Kinase B (*AURKB*), and Checkpoint Kinase 1 (*CHEK1*). Members of this family counter the expression of the tumor suppressor and cell cycle regulator Tumor Protein P53 (*TP53*), which is more highly expressed in healthy people. Moreover, all hub genes with higher transcript levels were associated with considerably poorer overall survival rates in BC patients. These results imply that these hub genes are relevant in terms of pathophysiology for the treatment of BC and deserve further attention. Kaplan-Meier survival analysis demonstrated that increased expression of all five genes was significantly associated with decreased survival ($P < 0.001$). Hazard ratios (HRs) ranged from 1.41 to 1.77, indicating a substantial negative impact on patient survival for each gene.

Conclusion: Survival analysis showed that tumors with higher expression levels of hub genes were associated with significantly shorter overall survival times among breast cancer patients. This finding suggests that these hub genes are highly relevant to BC pathophysiology and could be considered targets for monitoring.

Please cite this article as: Soleiman Ekhtiyari M, Ghaderi-Zefrehei M, Mogharari Z, Yousefi M, Bigdeli A, Nasre Esfahani E, Shahriarpour H, Leschm BJ. Integrated Expression Analysis May Support Serine/Threonine Kinases as Common Hub Genes in Breast Cancer. Iran J Med Sci. 2025;50(10):681-697. doi: 10.30476/ijms.2025.104386.3798.

Keywords • Breast neoplasms (D001943) • Computational biology (D057180) • Molecular modeling (D052199) • Gene regulatory networks (D059687) • Neoplasm proteins (D009369) • Protein-serine-threonine kinases (D051685) • Mortality (D009020)

Introduction

Breast cancer (BC) poses a significant threat worldwide.¹ A survey in 2020 revealed an estimated 276,480 new cases of BC diagnosed and approximately 42,170 deaths related to this disease in the United States alone.² Early detection and intervention can significantly hinder the disease progression to a metastatic state in numerous BC patients. However, a notable proportion, ranging from 20% to 30%, ultimately progresses to metastatic disease, which accounts for the majority of fatalities associated with breast cancer.^{2, 3} Various factors, including tumor subtype, response to initial treatment, and biological traits, are crucial in the onset of metastasis. Particularly aggressive subtypes, such as Receptor tyrosine-protein kinase *erbB-2* or Human Epidermal Growth Factor Receptor 2 (*HER2*)-positive, and triple-negative breast cancer, exhibit an increased susceptibility to metastasizing.^{4, 5} For early detection, precise prognosis, and developing specific treatments, it is essential to understand the genetic foundation underlying this malignancy.⁶ It is widely recognized that Breast cancer type 1 susceptibility protein 1 (*BRCA1*) and breast cancer type 2 susceptibility protein (*BRCA2*) mutations increase the risk of both breast and ovarian cancers.⁷⁻⁹ Mutations in other genes such as Ataxia-Telangiectasia Mutated (*ATM*), *BRCA1* Associated RING Domain 1 (*BARD1*), *BRCA1* Interacting Protein C-Terminal Helicase 1 (*BRIP1*), Caspase 8 (*CASP8*), Cytotoxic T-Lymphocyte Associated Protein 4 (*CTLA4*), Cytochrome P450 Family 19 Subfamily A Member 1 (*CYP19A*), Fibroblast Growth Factor Receptor (*FGFR*), Lymphocyte Specific Protein 1 (*LSP*), Mitogen-Activated Protein Kinase Kinase Kinase (*MAP3K*), Nibrin (*NBN*), RAD51 Recombinase (*RAD51*), and Telomerase Reverse Transcriptase (*TERT*) can also lead to an increased probability of developing BC.¹⁰⁻¹³ Most of these genes function as tumor suppressors and have a lower prevalence of mutations than *BRCA* genes and are generally associated with a lower risk factor for BC. The application of gene expression data in survival analysis among patients with BC has been explored in several studies.¹⁴⁻¹⁶ The *TP53* pathway, whose perturbation is a crucial factor in the development of BC, plays an essential role in maintaining genome stability. It does this by coordinating cell cycle arrest and apoptosis in response to DNA damage, thereby reducing the risk of passing on damaged genetic material.¹⁷

There are intricate connections between the proteins managing mitotic checkpoints and

regulating cell cycle progression, and those that contribute to genomic instability and tumor formation within the context of the *TP53* pathway.¹⁸ One intriguing aspect of this interplay is a tightly regulated feedback loop involving *TP53* and serine/threonine kinases, including WEE1 G2 Checkpoint Kinase (*WEE1*), Polo-Like Kinase 1 (*PLK1*), NIMA Related Kinase 2 (*NEK2*), BUB1 Mitotic Checkpoint Serine/Threonine Kinase (*BUB1*), TTK Protein Kinase (*TTK*), *AURKB*, and Aurora Kinase A (*AURKA*). The analysis of publicly available gene expression profiles has identified several kinases with potential synthetic lethal interactions with *TP53*, including *PLK1*, *NEK2*, *BUB1*, and *AURKA*. Therapeutic targeting of the catalytic domain of serine/threonine kinases is appealing because it offers a druggable target and can limit unfavorable effects on control cells. Disrupting essential signaling cascades by targeting kinases may also impede progression towards metastasis while treating BC effectively.¹⁹

BC remains a global health problem, with genetic factors such as germline *BRCA1/2* variants playing a key role in susceptibility. These mutations, which are widespread in different populations, play a role not only in breast and ovarian cancers but also in prostate cancer, as reported by Cioffi and colleagues.²⁰ Reproductive risk factors have a further influence on hereditary BC, as systematically analyzed by Springer.²¹ Advances in genetic modeling, such as the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) model validated by Møller and colleagues, improve forecasting pathogenic variants in cancer susceptibility genes and developing reliable tools for risk evaluation.²² Complementing these genetic insights, bioinformatic analyses, including Co-expression Network Analysis as applied by Xie and colleagues, have revealed disease mechanisms and important genetic factors.²³ Molecular studies, such as that of Asparuhova and colleagues, have disclosed the regulatory role of Transforming Growth Factor Beta (*TGF-β1*) and Insulin-Like Growth Factor 1 (*IGF-1*) in cancer progression and bone regeneration and identified potential therapeutic targets.^{24, 25} These integrated genetic, bioinformatic, and molecular approaches are crucial to unravel the complex mechanisms of BC and develop targeted therapies.

Innovative *in silico* techniques have facilitated investigations into gene functionality, diseases, and precision medicine at the molecular level.²⁶ Through an integrative approach to identify sets of genes associated with survival, Baculoviral IAP Repeat Containing 5 (*BIRC5*), Cyclin B1 (*CCNB1*),

and MYB Proto-Oncogene Like 2 (*MYBL2*) were found to have significant associations with BC patients' survival chances.¹⁵ Here, we aimed to conduct in-depth bioinformatic analyses of a large collection of publicly available datasets to identify hub genes and key pathways involved in BC. Our approach included examining the structural bioinformatics of these hub genes, with the ultimate goal of assessing their prognostic and therapeutic potential in BC.

Materials and Methods

Datasets

Following a keyword-driven search, we retrieved four gene expression data series from the NCBI GEO website (<https://www.ncbi.nlm.nih.gov/>). The datasets identified were GSE161533, GSE162228, GSE70947, and GSE139038. The GPL571 platform was utilized for generating datasets of both GSE161533 and GSE162228; whereas the dataset of GSE70947 was generated using the GPL13607 platform. The GPL27630 platform was used to create a microarray dataset for GSE139038. In total, the four datasets comprised 464 samples, including 248 BC samples and 216 control samples. Detailed information on the studies included in this analysis is provided in table 1.

To identify differentially expressed genes (DEGs) between cancerous and non-cancerous samples, the Bioconductor package was used. Perl (version 5.0) available online at <http://www.perl.org/> was employed for background correction as well as normalization purposes afterwards. We retrieved an equivalent set of RNA-Seq datasets by preserving the same keyword search as above. Data was downloaded from TCGA (The Cancer Genome Atlas) (<https://cancergenome.nih.gov/>), and analyzed using the limma package (an R add-on) using cut-off criteria of adjusted $P < 0.05$ and $|\log_2FC| > 1$. This dataset comprised 563 BC samples and 114 control samples in total, as obtained from TCGA's BC dataset. For gene expression calculations, the edgeR package was used, following similar methods used to evaluate integrated GEO datasets for consistency in data analysis. Figure 1 depicts a flowchart outlining the data analysis pipeline.

Weighted Gene Co-expression Network Analysis

The expression data profiles were used to generate gene co-expression networks using the Weighted Gene Co-expression Network Analysis (WGCNA) package in R (version 1.68). A weighted adjacency matrix was constructed using the power function β , as previously described.^{31, 32} We selected an appropriate β value to enhance matrix similarity and construct a co-expression network. Hierarchical weighting matrix clustering was used to define the modules. Intermediate graft hierarchies were conducted based on the calculation of Topological Overlap Matrix (TOM) -based dissimilarity, with the lowest size considered for the gene dendrogram. Genes with similar expression indices were categorized within identical gene units using the Dynamic Tree Cut algorithm. The Enrichr online database (<https://maayanlab.cloud/Enrichr/>) was used to analyze the molecular and functional characteristics and KEGG pathways of the DEGs. Statistical significance was set at $P < 0.05$. The STRING database (<https://string-db.org/>) was applied to obtain protein-protein interaction (PPI) information for the DEGs. Then, Cytoscape software (<https://cytoscape.org/>) (version 3.9.1) was employed to assemble a PPI association network.

Table 1: Detailed information on the studies included in this analysis.

GSE No.	Study	Number of samples	Platform	Organism
GSE161533	Expression data from esophageal squamous cell carcinoma patients ²⁷	84	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	<i>Homo sapiens</i>
GSE162228	Concordance of PAM50 molecular subtyping between oligonucleotide microarray and NanoString nCounter assay for Taiwanese BC ²⁸	133	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array	<i>Homo sapiens</i>
GSE70947	Age and estrogen-dependent inflammation in breast adenocarcinoma and normal breast tissue [cohort_2] ²⁹	296	Agilent-028004 SurePrint G3 Human GE 8x60K Microarray (Feature Number version)	<i>Homo sapiens</i>
GSE139038	Gene expression profiling in paired normal, apparently normal, and breast tumor tissues ³⁰	65	Print_1437(Block_Column_Row IDs)	<i>Homo sapiens</i>

[HG-U133-Plus-2] Affymetrix Human Genome U133 Plus 2.0 Array: A widely used microarray platform for analyzing gene expression profiles. It covers over 47,000 transcripts and variants, providing comprehensive coverage of the human genome; Agilent-028004 SurePrint G3 Human GE 8x60K Microarray (version with function number): A high-resolution microarray platform with 60,000 probes per array for detailed gene expression analysis with increased sensitivity and specificity in transcript detection; Print_1437 (Block-Column-Row-IDs): A custom microarray platform organized by block, column, and row identifiers to uniquely identify probes and enable precise measurement of gene expression.

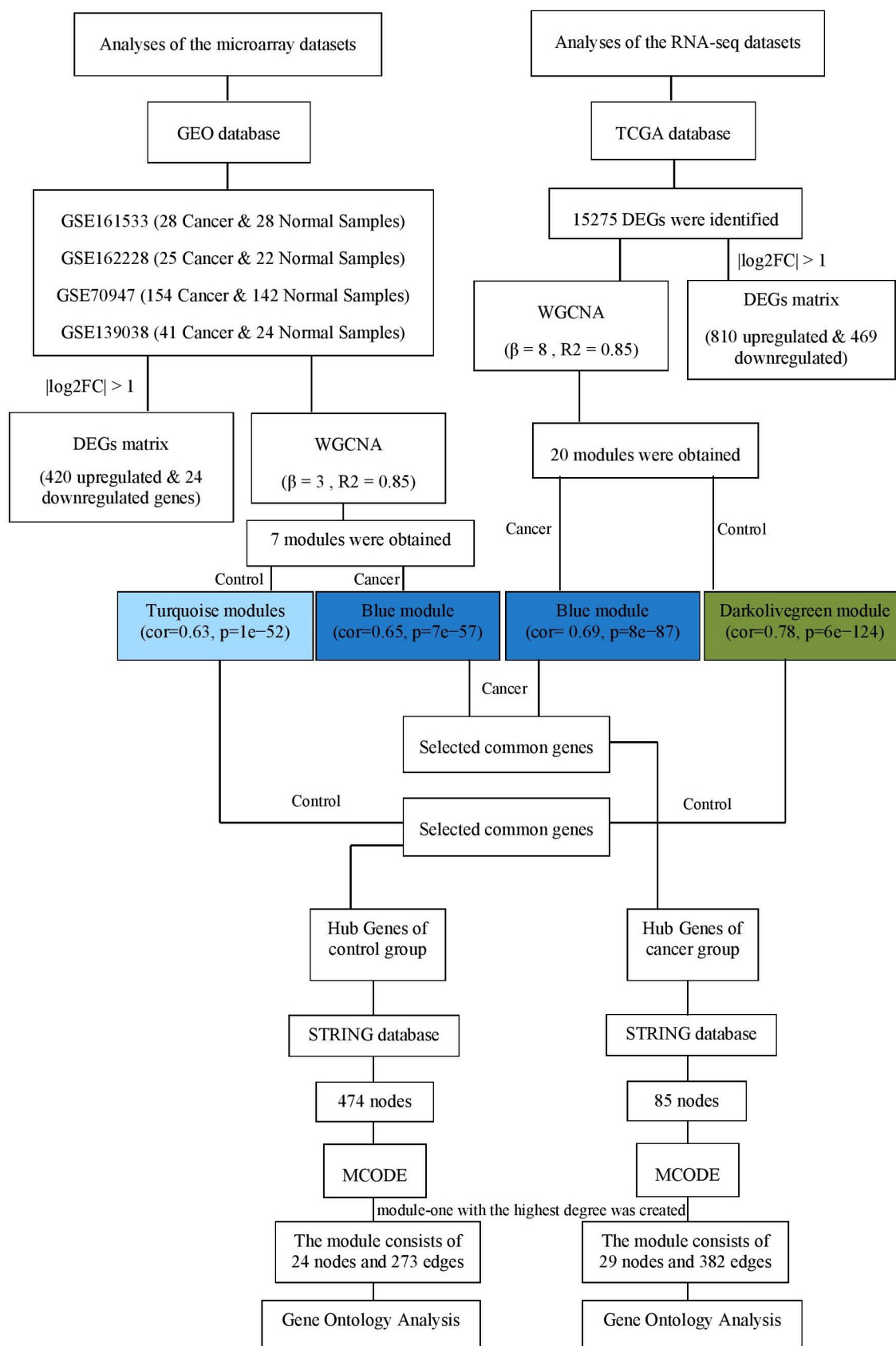


Figure 1: This figure illustrates the comprehensive bioinformatics data analysis pipeline used here for analyzing gene expression data from both microarray and RNA-seq datasets, sourced from the GEO and TCGA databases. The process begins with the retrieval and preprocessing of four datasets from the GEO database and RNA-seq data from TCGA, comprising a total of 464 samples (248 breast cancer samples and 216 control samples) for GEO and 677 samples (563 breast cancer samples and 114 control samples) for TCGA.

The Molecular Complex Detection (MCODE) plug-in in Cytoscape was employed to extract core gene modules from the PPI network, applying the following parameters: degree cut-off ≥ 2 , node score cut-off ≥ 0.2 , K-core ≥ 2 , and a maximum depth of 100.

Statistical Analysis

The Kaplan-Meier Plotter online database (<http://www.kmplot.com>) was employed to perform the survival analyses of the selected core genes. The hazard ratio (HR) with 95% confidence intervals and log-rank P value were determined and shown on the plot.

Structural Bioinformatics for Upregulated Serine/Threonine Kinases

To acquire insights into the functionality of the serine/threonine kinase protein family, we employed a robust methodology that combines both sequence and structural alignments. Sequencing and structural alignments for these

serine/threonine kinase proteins were conducted using the YASARA (version 19.12.14) software and the PRALINE service with the MUSTANG algorithm.^{33, 34} Our working hypothesis centers on the idea that, when analyzed in the context of protein structures, structural and sequence alignments may provide mechanisms for controlling the activity of this family of proteins.

Results

Determination of DEGs in BC

Analysis of BC samples in comparison to control samples of the four GEO microarray data series (GSE161533, GSE162228, GSE70947, and GSE139038) revealed upregulated and downregulated genes for each of the selected datasets (GSE161533: 476 upregulated, 399 downregulated; GSE162228: 183 upregulated, 109 downregulated; GSE70947: 285 upregulated, 292 downregulated; GSE139038: 1016 upregulated, 391 downregulated) (figure 2).

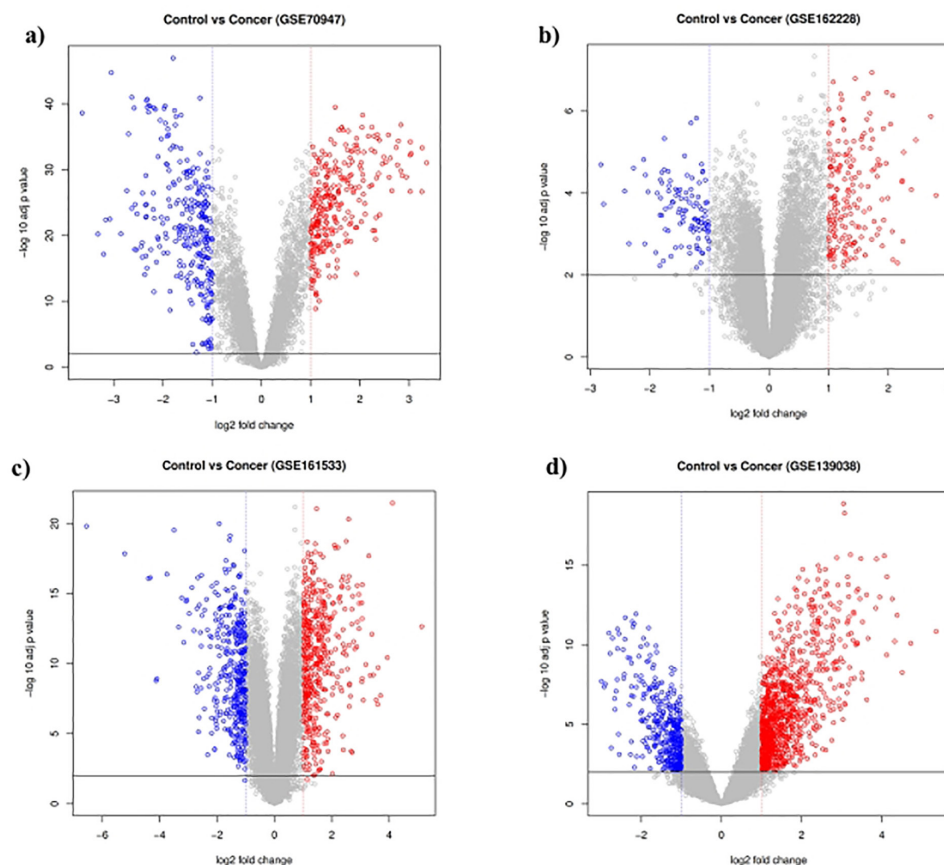


Figure 2: Volcano plots a, b, c, and d are presented for the GSE70947, GSE162228, GSE161533, and GSE139038 datasets. In general, this figure depicts the DEGs between control and BC samples across four distinct datasets (GSE70947, GSE162228, GSE161533, and GSE139038). Each plot features the log2 fold change (log2FC) on the horizontal axis and the -log10 P value on the vertical axis. Genes that are significantly upregulated are marked in red (log2FC >1 , $P<0.05$), while downregulated genes are indicated in blue (log2FC <-1 , $P<0.05$), and non-significant genes are shown in gray. The plots distinctly separate significantly DEGs from those that are not significant, underscoring the influence of BC on gene expression. The variations in log2FC and -log10 P values across the datasets reflect differences in the extent and significance of gene expression alterations. These visual representations are essential for pinpointing critical genes associated with cancer biology, which may act as potential biomarkers or therapeutic targets for future research.

In other words, figure 2 shows volcano plots of differentially expressed genes (DEGs) from four datasets (GSE70947, GSE162228, GSE161533, GSE139038) comparing BC (BC) and control samples ($P < 0.05$, $|\log_2FC| > 1$). The x-axis represents the $\log_2$ fold changes (extent of gene expression), and the y-axis shows the $-\log_{10}$ P values (statistical significance). Red and blue dots indicate significantly upregulated and downregulated genes in BC, respectively, while gray dots represent non-significant genes. The graphs highlight important DEGs in different datasets and show potential biomarkers and therapeutic targets for BC.

The analysis of the four integrated datasets together identified a total of 1083 shared genes, with 444 DEGs, comprising 420 upregulated and 24 downregulated genes in BC samples compared to control samples ($|\log_2FC| > 1$). There were 15275 genes found in the RNA-Seq dataset (brca-tcga-pan-can-atlas-2018) from the TCGA database. The analysis of this dataset identified 1279 DEGs in total, including 810 upregulated and 469 downregulated genes in BC samples compared to control samples ($|\log_2FC| > 1$). Principal components analysis (PCA) of the datasets revealed that they were of sufficient quality for further bioinformatics processing. The PCA of expressed genes showed clustering of the expression profiles of control compared to BC samples, as well as adequate quality evaluation markers for future

data processing (figure 3). Figure 3 shows plots of PCA for (a) the integrated microarray dataset and (b) the TCGA dataset, showing a clear separation between control (con) and cancer (col) groups. Each point represents a sample, with the first principal component 1 (PC1) and the second principal component 2 (PC2) showing the highest variance. The clear clustering of the groups emphasizes the different expression profiles between cancer and control samples and confirms the quality of the dataset and the consistency of the grouping. Weighted correlation network analysis found that similar genes from two datasets were consistently expressed in BC samples, in contrast to the control samples.

WGCNA Microarray Datasets

We analyzed 2335 genes with comparable expression levels using the WGCNA R program. To determine if the network was scale-free, we set the power of ≈ 3 (scale-free $R^2 = 8.5$) (figures 4a and 4b) as a soft-threshold parameter. The modulus component connection is depicted in figure 5a in this regard. The modules' dissimilarity was set at 0.25, and a total of 7 modules were produced. A correlation heat-map shows a set of modules (turquoise, $\text{cor} = 0.63$, $P = 1e52$) that were near to each other in the control samples and a distinct set of modules (blue, $\text{cor} = 0.65$, $P = 7e57$) near each other in the cancer samples (figure 5b).

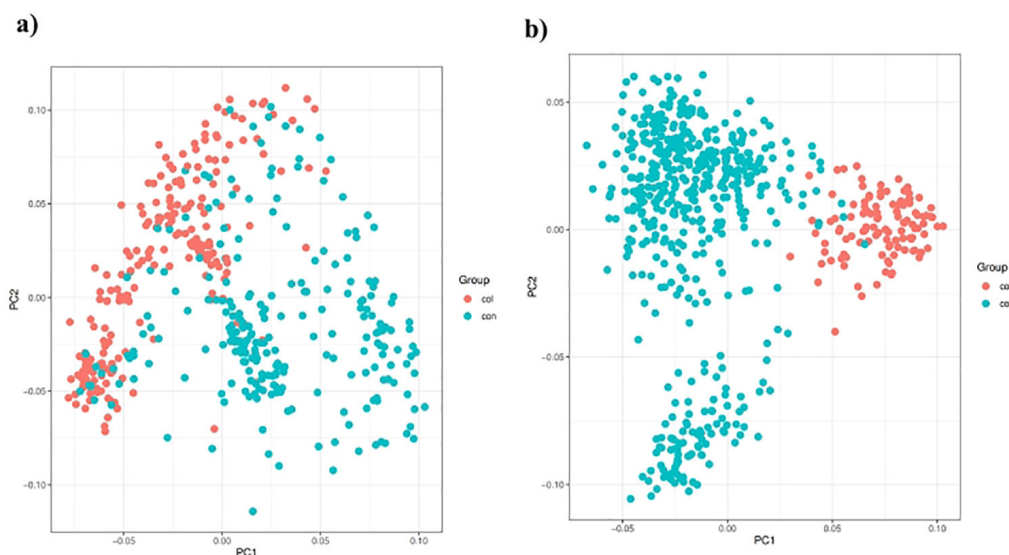


Figure 3: Principal Component Analysis (PCA)-based evidence of (a) integrated microarray and (b) TCGA dataset. To this end, figure 3 illustrates the results of PCA for (a) a combined microarray dataset and (b) the TCGA dataset, demonstrating the differentiation between control (con) and BC (can) samples. Each data point corresponds to a sample, with color coding employed to distinguish the groups: red denotes cancer samples, while blue represents control samples. The PCA visualizations depict the first two principal components (PC1 and PC2), which account for the majority of the variance present in the datasets. In both cases, a clear clustering of cancer and control samples is observed, signifying distinct gene expression patterns between the two categories. This observed separation indicates that PCA is effective in reducing data dimensionality while maintaining the variance that differentiates cancer from control samples, underscoring the utility of these datasets in uncovering critical molecular distinctions related to BC.

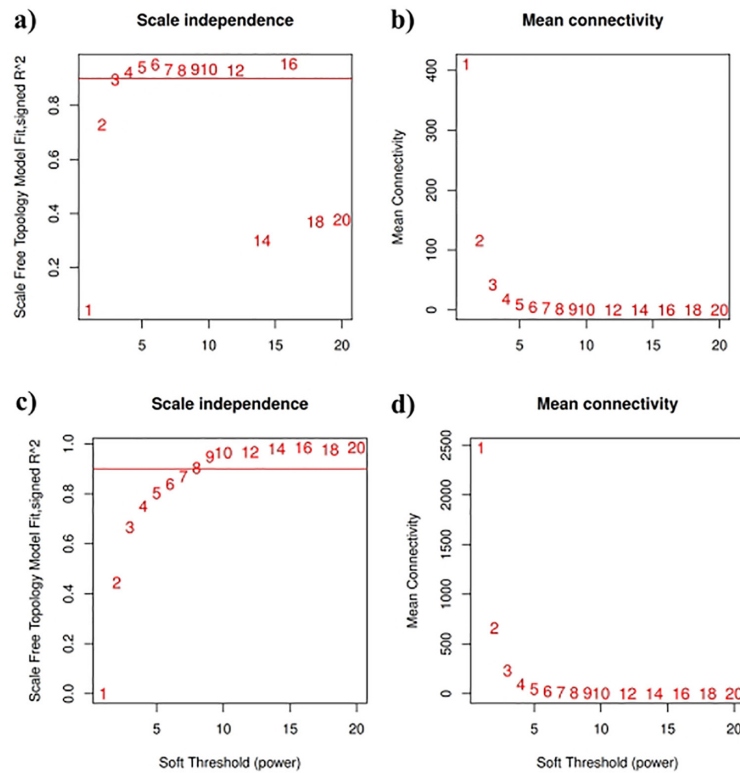


Figure 4: The figure shows the selection of soft-thresholding power (β) for microarray (a, b) and RNA-Seq datasets (c, d). The selected β ensures an $R^2 \geq 0.85$ for the scale-free topology (a, c) while maintaining a reduced mean connectivity (b, d), which optimizes the network design.

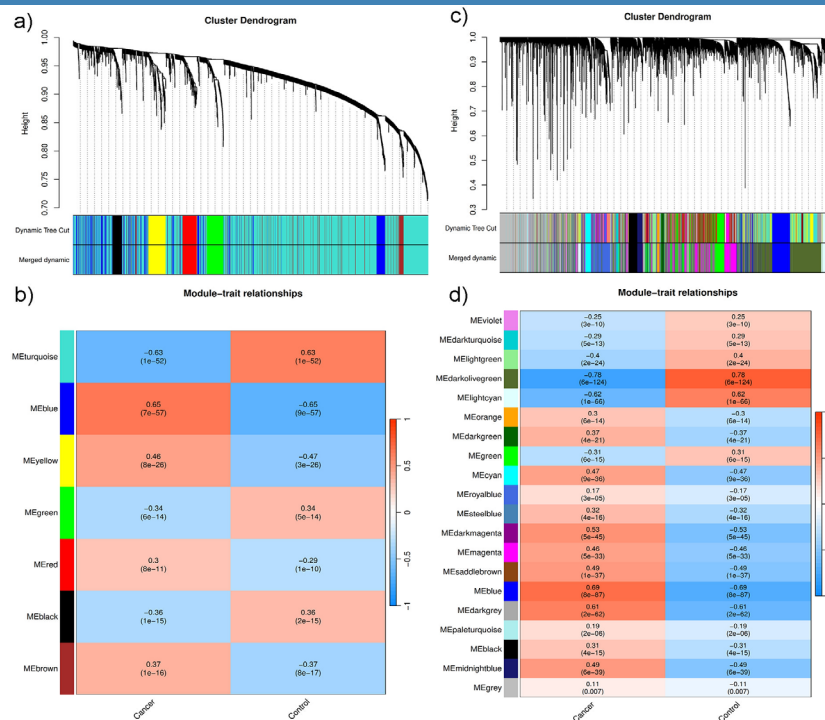


Figure 5: This figure illustrates WGCNA of gene expression data about BC, contrasting microarray datasets (a, b) with RNA-Seq datasets (c, d) to uncover co-expressed gene modules and their relationships with BC phenotypes. The dendrograms (a, c) depict the hierarchical clustering of genes into distinct modules, represented by color-coded branches. While clustering patterns exhibit similarities across all datasets, some variations are also evident. In the module-trait relationships (b, d), the heatmaps display the correlations between module eigengenes (MEs) and BC samples. Notably, significant modules identified in both datasets imply a potential involvement in cancer development or progression, although the correlation strength may differ due to variations in technology. Larger modules could signify fundamental processes, whereas smaller modules might reflect specialized functions. Gene ontology enrichment analysis may uncover biological pathways linked to BC. Comparing these results with findings from other studies can further validate the conclusions and provide new insights. This analysis underscores critical gene modules that may play a role in BC, presenting opportunities for the identification of therapies and biomarkers.

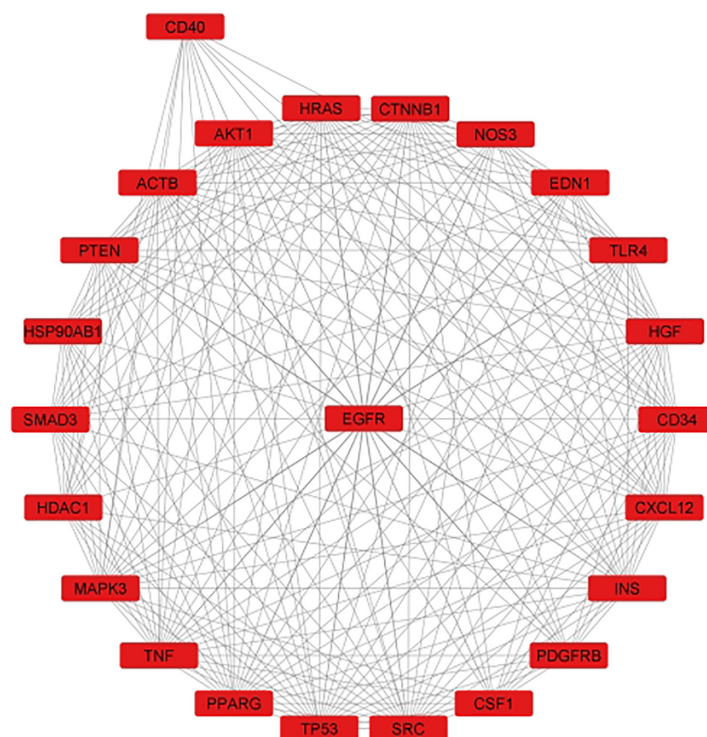


Figure 6: The protein-protein interaction (PPI) network is illustrated in control samples. This figure represents protein interactions observed in control samples, emphasizing significant characteristics such as a high density of nodes, which signifies intricate interactions, and central hub proteins such as Epidermal Growth Factor Receptor (*EGFR*) and CD40 (*CD40*), which act as key regulators. The modular architecture of the network implies the existence of functional complexes or pathways, while the discovery of novel interactions may uncover previously unrecognized connections. Important considerations include the necessity for experimental validation to reduce false positives, employing gene ontology and literature reviews to assess functional significance, and acknowledging the dynamic nature of PPI networks, which fluctuate according to cellular conditions.

RNA-Seq Datasets

To determine if the network was scale-free, we set the power of=8 (scale-free $R^2=8.5$) (figures 4c and 4d) as a soft-threshold parameter. The modulus component connection is depicted in figure 5c in this regard. The dissimilarity threshold for the modules was set to 0.25, resulting in the generation of 20 distinct modules. Similar to the results from microarray data, a specific module set (blue, figure 5d) was grouped in BC samples, while a distinct module set (dark olive green, figure 5d) was grouped in the control samples.

Selection of Hub Genes and Pathway Analysis

Control Samples: The common genes of the dark-olive-green (control samples from microarray datasets) and turquoise (control samples from RNA-Seq datasets) modules were combined to define a set of genes whose association with control compared to tumor tissue was well supported. Then, based on the amount of interaction, Cytoscape and the online database STRING were used to choose key genes and significant gene modules. 474 DEGs were categorized into a DEG-based PPI network complex after calculations were done. MCODE analysis was applied to identify specific PPI network components, and a module with

the highest degree was generated. The module consists of 24 nodes and 273 edges (figure 6).

BC Samples: The common genes of the blue (BC-associated) modules from microarray and RNA-Seq datasets were combined to define a set of genes whose association with tumor compared to control tissue was well supported. Cytoscape and STRING were used to select the fundamental genes and significant gene modules linked to BC based on the amount of interaction. A PPI network was created (figure 7), and MCODE was applied to identify specific modules within the PPI network. The module with the highest degree included 382 edges and 29 nodes (figure 7).

Hub Gene Ontology Analysis in BC

Control Samples: To learn more about the functional properties of the first module, Gene Ontology (GO) analysis was performed using the Enrichr online database (<https://maayanlab.cloud/Enrichr/>). The first control tissue-associated module was analyzed according to three functional classes, namely Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). The top 10 classes were reported according to P value. Hub genes were associated with specific enrichments in each of the three groups.

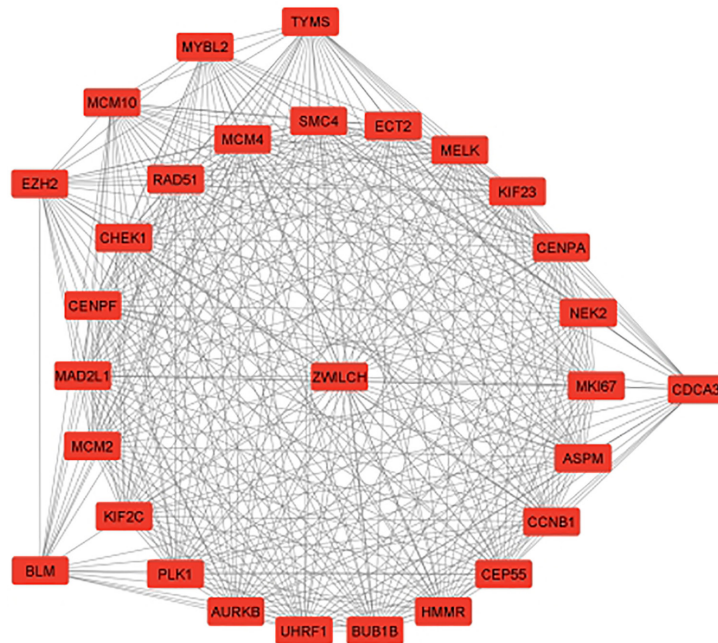


Figure 7: The protein-protein interaction (PPI) network is illustrated in BC samples. This figure emphasizes the interactions among proteins and their functional interrelations within the context of BC. Notable characteristics include a high density of nodes, indicating a complex network of protein interactions, as well as the presence of hub proteins such as ZWINT, TYMS, and MCM10, which are pivotal in orchestrating the network. The modular architecture of the network suggests the existence of functional protein complexes or pathways pertinent to BC, including several novel interactions that have not been documented previously. It is crucial to consider the validation of these interactions, the exploration of their functional significance through gene ontology, and the recognition of the dynamic characteristics of PPI networks.

Positive regulation of cellular processes, positive regulation of cellular biosynthetic processes, positive regulation of protein phosphorylation, positive regulation of protein serine/threonine kinase activity, and regulation of cell population proliferation were the main functions of hub genes in the BP class. In the CC class, hub genes were particularly enriched in focal adhesion, cell-substrate junction, vesicle, membrane

raft, caveola, and bounding membrane of organelle. Finally, for the MF class, hub genes were mostly enriched for RNA polymerase II-specific DNA-binding transcription factor binding, ubiquitin protein ligase binding, DNA-binding transcription factor binding, ubiquitin-like protein ligase binding, transcription coregulator binding (figure 8). KEGG pathway enrichment analysis was also performed for the hub genes.

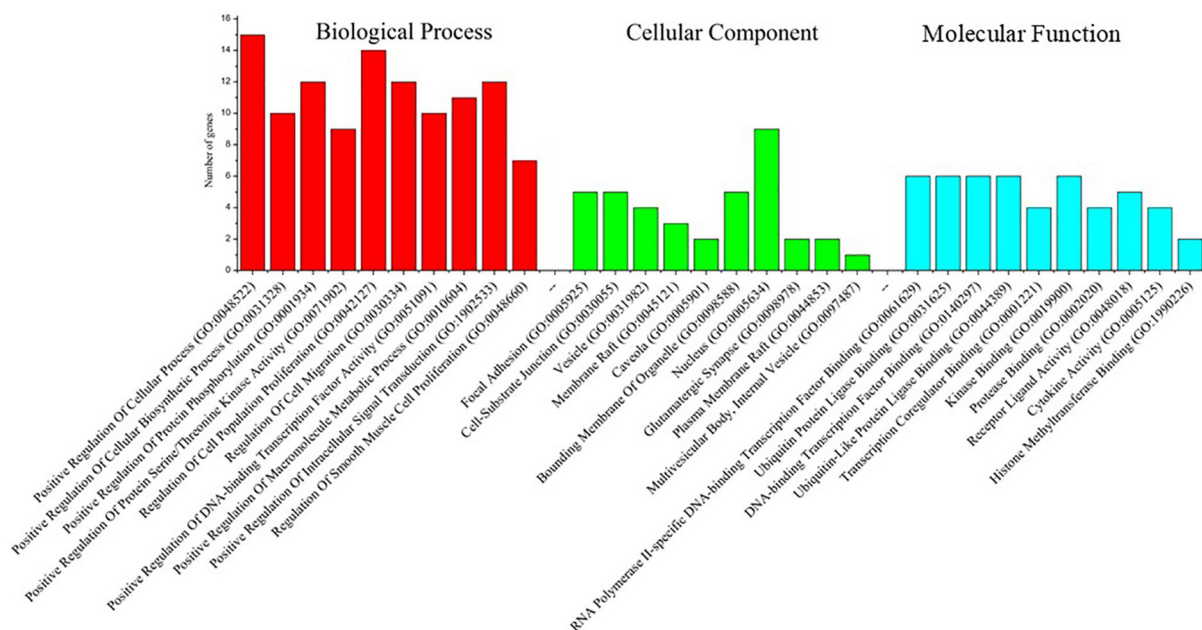


Figure 8: The Gene Ontology analysis of hub genes shows upregulation in control samples compared to BC samples.

Table 2: Significantly enriched examination of the identification of differentially expressed genes in control samples in the KEGG pathway

Term	P value	Adj. P value	Genes
Cell cycle	4.14E-10	7.86E-09	<i>CCNB1; CHEK1; PLK1; BUB1B; MCM4; MAD2L1; MCM2</i>
Progesterone-mediated oocyte maturation	4.03E-04	0.0038306	<i>CCNB1; PLK1; MAD2L1</i>
Oocyte meiosis	8.47E-04	0.005089732	<i>CCNB1; PLK1; MAD2L1</i>
DNA replication	0.0012404	0.005089732	<i>MCM4; MCM2</i>
Cellular senescence	0.0014656	0.005089732	<i>CCNB1; CHEK1; MYBL2</i>
Homologous recombination	0.0016073	0.005089732	<i>BLM; RAD51</i>
Fanconi anemia pathway	0.0027724	0.007525035	<i>BLM; RAD51</i>
Human T-cell leukemia virus 1 infection	0.0038357	0.009109689	<i>CHEK1; BUB1B; MAD2L1</i>
p53 signaling pathway	0.0050055	0.010567085	<i>CCNB1; CHEK1</i>
FoxO signaling pathway	0.0153991	0.029258385	<i>CCNB1; PLK1</i>

The top four enriched KEGG pathways were cell cycle, progesterone-mediated oocyte maturation, oocyte meiosis, and DNA replication. The list of DEGs (top 10 according to P value) for each pathway is listed in table 2. To give a better picture to the reader, figure 8 illustrates a GO analysis that depicts the distribution of hub genes across different GO categories; namely, BP, CC, and MFs were identified as being upregulated in control samples compared to BC samples. Notable findings indicate that a majority of the hub genes participate in biological processes such as cellular regulation and protein phosphorylation, are associated with cellular components such as focal adhesions and vesicles, and perform functions including transcription factor binding and receptor-ligand interactions. The increased

expression of these genes in control samples implies their possible role as tumor suppressors in BC. It is crucial to further analyze and validate their functional significance, as GO analysis only provides a general perspective on gene functions.

Cancer Samples: The genes with the smallest enrichment P value in the cancer hub study were those associated with “Single-Stranded DNA Binding” in MF, “Regulation of Cell Cycle Process” in BP, and “Spindle” in CC (figure 9). Figure 9 presents the results of GO enrichment analysis about BC-associated genes, categorized into three distinct domains: BP, represented in red, CC, depicted in green, and MF, illustrated in blue. The x-axis enumerates the enriched GO terms, while the y-axis reflects the corresponding gene counts.

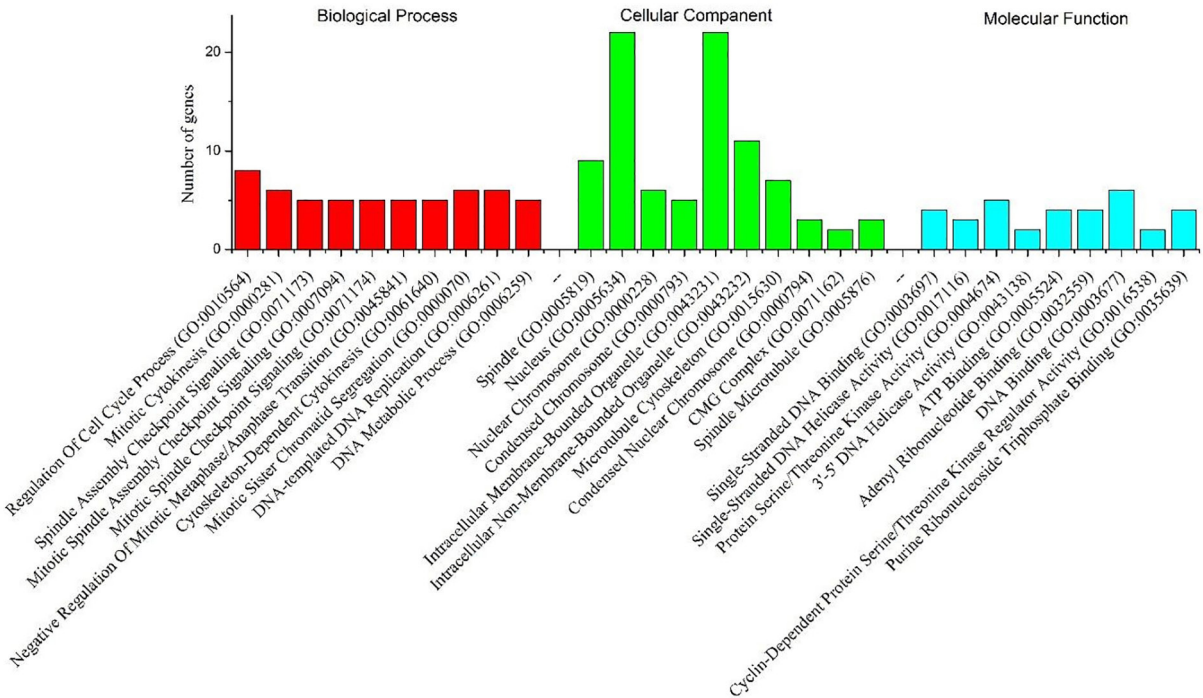


Figure 9: Gene Ontology analysis of Identification of Differentially Expressed Genes in BC showed DEGs were organized within three functional classes, namely biological process (BP), cellular component (CC), and molecular function (MF). The top 10 significantly enriched GO terms in BC are shown for each class.

Table 3: The enrichment of DEGs in BC in the KEGG pathway (Top 15 according to P value)

Term	P value	Adj. P value	Genes
Pathways in cancer	1.12E-16	2.11E-14	<i>PDGFRB</i> ; <i>EDN1</i> ; <i>SMAD3</i> ; <i>HSP90AB1</i> ; <i>HDAC1</i> ; <i>HGF</i> ; <i>PTEN</i> ; <i>EGFR</i> ; <i>CXCL12</i> ; <i>AKT1</i> ; <i>CTNNB1</i> ; <i>PPARG</i> ; <i>TP53</i> ; <i>HRAS</i>
Prostate cancer	1.25E-15	1.10E-13	<i>PDGFRB</i> ; <i>HSP90AB1</i> ; <i>PTEN</i> ; <i>AKT1</i> ; <i>CTNNB1</i> ; <i>TP53</i> ; <i>HRAS</i> ; <i>EGFR</i> ; <i>INS</i>
PI3K-Akt signaling pathway	1.75E-15	1.10E-13	<i>PDGFRB</i> ; <i>HSP90AB1</i> ; <i>CSF1</i> ; <i>NOS3</i> ; <i>HGF</i> ; <i>PTEN</i> ; <i>AKT1</i> ; <i>TP53</i> ; <i>HRAS</i> ; <i>TLR4</i> ; <i>EGFR</i> ; <i>INS</i>
Proteoglycans in cancer	1.78E-14	8.35E-13	<i>SRC</i> ; <i>HGF</i> ; <i>AKT1</i> ; <i>CTNNB1</i> ; <i>TNF</i> ; <i>TP53</i> ; <i>HRAS</i> ; <i>TLR4</i> ; <i>EGFR</i> ; <i>ACTB</i>
Rap1 signaling pathway	2.27E-14	8.52E-13	<i>PDGFRB</i> ; <i>CSF1</i> ; <i>SRC</i> ; <i>HGF</i> ; <i>AKT1</i> ; <i>CTNNB1</i> ; <i>HRAS</i> ; <i>EGFR</i> ; <i>ACTB</i> ; <i>INS</i>
Lipid and atherosclerosis	2.87E-14	9.00E-13	<i>CD40</i> ; <i>HSP90AB1</i> ; <i>SRC</i> ; <i>NOS3</i> ; <i>AKT1</i> ; <i>PPARG</i> ; <i>TNF</i> ; <i>TP53</i> ; <i>HRAS</i> ; <i>TLR4</i>
Fluid shear stress and atherosclerosis	3.48E-14	9.35E-13	<i>EDN1</i> ; <i>HSP90AB1</i> ; <i>SRC</i> ; <i>NOS3</i> ; <i>AKT1</i> ; <i>CTNNB1</i> ; <i>TNF</i> ; <i>TP53</i> ; <i>ACTB</i>
Hepatocellular carcinoma	1.97E-13	4.63E-12	<i>SMAD3</i> ; <i>HGF</i> ; <i>PTEN</i> ; <i>AKT1</i> ; <i>CTNNB1</i> ; <i>TP53</i> ; <i>HRAS</i> ; <i>EGFR</i> ; <i>ACTB</i>
Focal adhesion	1.00E-12	2.09E-11	<i>PDGFRB</i> ; <i>SRC</i> ; <i>HGF</i> ; <i>PTEN</i> ; <i>AKT1</i> ; <i>CTNNB1</i> ; <i>HRAS</i> ; <i>EGFR</i> ; <i>ACTB</i>
Melanoma	1.91E-12	3.60E-11	<i>PDGFRB</i> ; <i>HGF</i> ; <i>PTEN</i> ; <i>AKT1</i> ; <i>TP53</i> ; <i>HRAS</i> ; <i>EGFR</i>

Notable observations include “Regulation of Cell Cycle Process” within the BP domain, “Spindle” in the CC domain, and “Single-Stranded DNA Binding” in the MF domain, all of which underscore their contributions to cancer progression. These terms highlight essential processes involved in the regulation of the cell cycle, the functionality of the mitotic spindle, and the mechanisms of DNA replication and repair, thereby emphasizing their critical roles in maintaining genomic stability and facilitating cancer development.

Based on KEGG pathway enrichment in the cancer hub genes, the top four enriched pathways were pathways in cancer, prostate cancer, PI3K-Akt signaling pathway, and proteoglycans in cancer. In table 3, the top 10 DEGs for each enriched category are shown in order of P values. Five genes, including *NEK2*, *MELK*, *PLK1*, *AURKB* (or *STK12*), and Checkpoint Kinase 1 (*CHEK1*), which are in the kinase enzyme family and known to oppose *TP53* function, were found during this analysis.

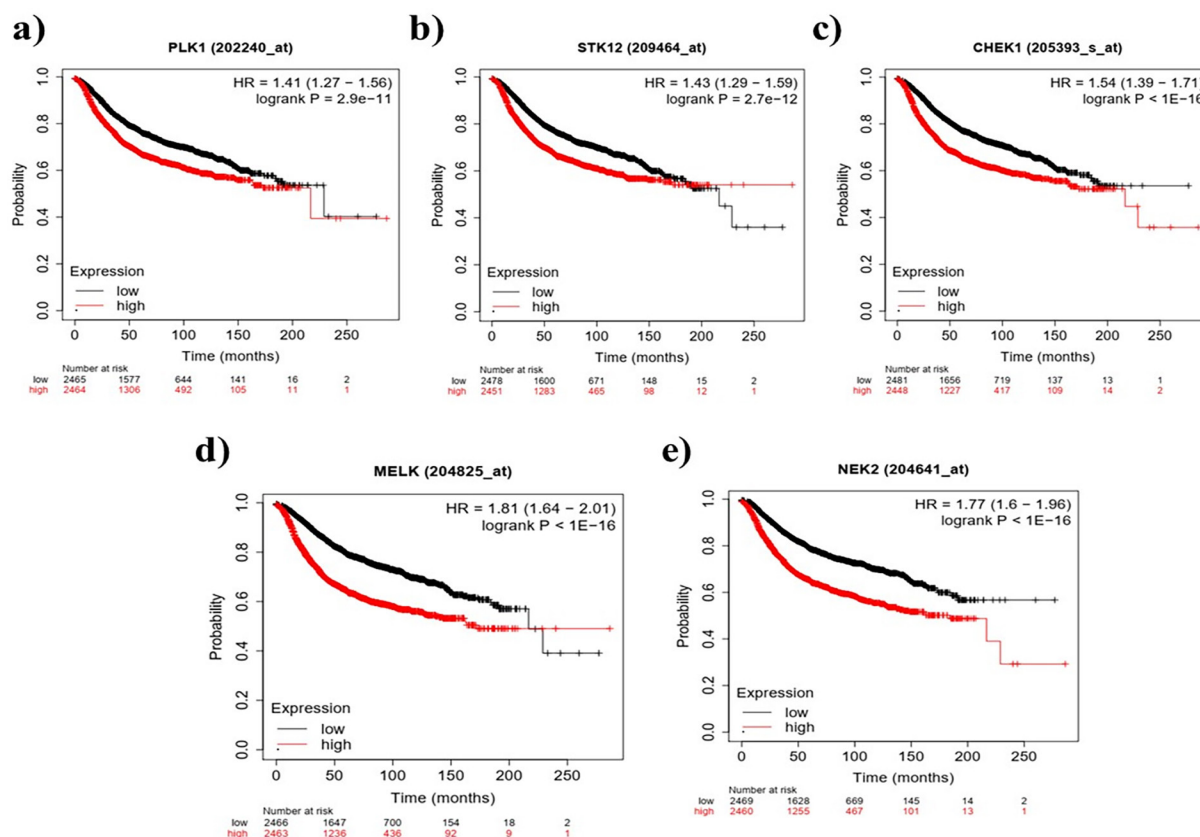


Figure 10: The prognostic values of the nine core genes in BC are shown. (a) *PLK1*; (b) *AURKB* (*STK12*); (c) *CHEK1*; (d) *MELK*; (e) *NEK2*. This figure depicts Kaplan-Meier survival analyses for five genes associated with BC: *PLK1*, *STK12*, *CHEK1*, *MELK*, and *NEK2*. It compares the survival probabilities of groups with high expression (represented by red curves) against those with low expression (indicated by black curves). The analysis reveals that elevated expression levels of all five genes correlate significantly with reduced survival rates, as evidenced by hazard ratios (HRs) ranging from 1.41 to 1.77 and log-rank P values below 0.001.

Survival Analysis of Hub Genes

We carried out survival analysis to investigate the relationship between gene expression of each of the hub genes and the overall survival time for BC patients to better understand the predictive significance of hub genes identified in this study and uncover possible cancer progression-related protein-encoding genes. We concentrated on five important kinase genes (*NEK2*, *MELK*, *PLK1*, *AURKB* (*STK12*), and *CHEK1*) that affect pathways linked with BC development. Depending on their median expression level, patients were divided into groups with high or low expression. All hub genes with higher transcript levels were associated with considerably poorer overall survival rates in BC patients. These results imply that these hub genes are relevant in terms of pathophysiology for the treatment of BC and deserve further attention (figure 10).

Structural Aspect of Serine/Threonine Kinase

The serine/threonine kinases identified as hub genes (*NEK2*, *MELK*, *PLK1*, *AURKB* (*STK12*), and *CHEK1*) were significantly similar in their amino acid sequences as demonstrated

by the protein structure and sequence alignment shown in figure 11. These kinases therefore share commonalities in conserved regions and crucial amino acids, particularly within their binding areas. Therefore, future studies may focus on the high-similarity sites in this family, which are involved in both binding and activation processes, to develop targeted BC treatments. These findings are broadly applicable beyond this study as members of other families may share similar structural features.

Discussion

BC remains a significant challenge in oncology due to its complex pathophysiology and resistance to treatments. This study identified a high degree of structural similarity in the binding and active site domains of these five kinases (*NEK2*, *MELK*, *PLK1*, *AURKB* (*STK12*), and *CHEK1*), suggesting the potential for developing therapeutic strategies targeting all five simultaneously. These kinases are known to play roles in tumorigenesis, including immune cell infiltration, immune escape, cell proliferation, and cell cycle regulation. In this study, we

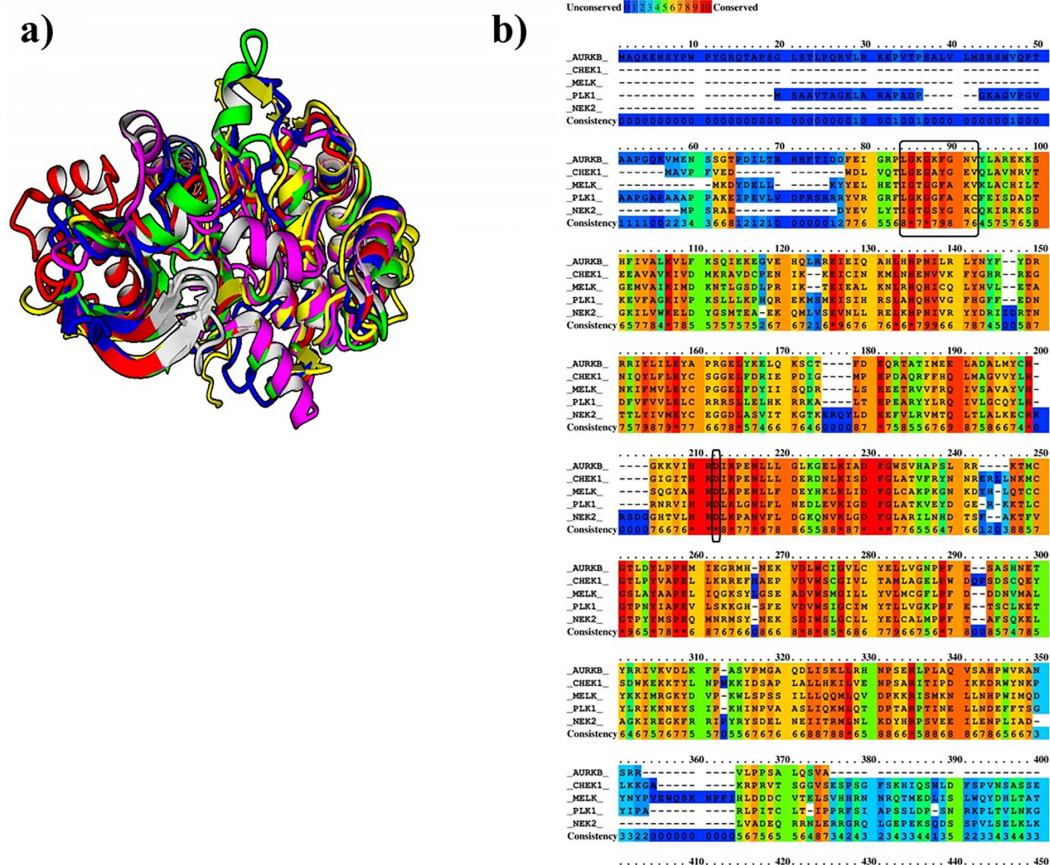


Figure 11: This figure shows (a) 3D structural alignment of serine/threonine kinase proteins. Alignment is shown with the protein-binding site highlighted in gray. The amino acids of this motif (gray) are specifically indicated in the box in the b section. (b) Alignment of serine/threonine protein sequences. The binding site is highlighted in the first box (LGKGFNGV, LGEGAYGEV, IGTGGFAKV, LGKGGFAKC, IGTGSYGRC), and the common active site "D" is highlighted in the second box.

conducted a large-scale transcriptomic analysis integrating microarray and RNA-seq datasets, which identified 444 DEGs. Most DEGs (420) were up-regulated in BC, while 24 were down-regulated. Through co-expression network analysis, two distinct gene expression modules were identified: one enriched in normal tissue (dark olive green) and another in BC tissue (blue). PPI network analysis further identified 29 hub genes in BC, of which five serine/threonine kinase genes (*NEK2*, *MELK*, *PLK1*, *AURKB*, and *CHEK1*) stood out due to their significant roles in tumor progression and potential as therapeutic targets. We performed additional analysis for these five hub genes, which are known to functionally oppose TP53's role in maintaining cellular homeostasis and have potential for therapeutic use in BC. We found a high degree of commonality in the structures of the binding and active site domains among these five proteins. This observation suggests that future research endeavors could explore therapeutic methods to simultaneously regulate the expression or function of these proteins via their highly conserved domains.

Each of the identified five kinase proteins has been previously linked to tumorigenesis. *MELK* is closely connected to immune cell infiltration, function, and production of cytokines such as Interferon Gamma (IFN- γ) and Interferon Alpha (IFN- α) in BC.³⁵ The *NEK2* may cause tumor immune escape by altering Programmed Death-Ligand 1 (*PD-L1*) expression in multiple cancers, including BC.^{36, 37} High expression of the *PLK1* gene has also frequently been observed in breast and other cancers, and has been targeted as a potential therapeutic using both siRNA depletion and small-molecule inhibitors. Reducing *PLK1* activity can effectively restrain proliferation and promote apoptosis among tumor cells.³⁸ *AURKB* is overexpressed in numerous human tumors and regulates multiple processes, including chromatin condensation and the phosphorylation of histone H3 on Ser10 and centrosome type A protein on Ser7 in early G2.³⁹ Finally, conserved *CHEK1* plays an essential rate-limiting role during the cell cycle, and its overexpression can lead to tumorigenesis.⁴⁰ Together, these studies suggest that the five selected hub genes are functionally connected and share structural similarities. This reinforces our finding that these genes are key hubs in BC, highlighting their critical role.

Serine/threonine kinases are integral to the advancement and metastasis of BC, acting as essential regulators of cellular signaling pathways that facilitate tumor growth and spread. Their aberrant regulation has been associated

with numerous oncogenic mechanisms, thereby establishing these kinases as attractive targets for therapeutic intervention. Rupasinghe and colleagues emphasized their importance in cancer biology, pointing out the potential of small-molecule inhibitors to impede tumor progression.⁴¹ Ghafouri-Fard and colleagues concentrated on cyclin-dependent kinases, a specific category of serine/threonine kinases, and demonstrated their crucial involvement in cell cycle dysregulation and the pathogenesis of cancer.⁴² Saavedra highlighted the therapeutic potential of mitotic kinases, especially in the context of aggressive forms such as triple-negative BC.⁴³ Additional research has identified these kinases as central genes in tumor dynamics, reinforcing their significant role in cancer biology and progression. Furthermore, advancements in structural bioinformatics have facilitated the discovery of druggable sites within serine/threonine kinases, which could lead to the development of targeted therapies.

Our study underscores the significance of analyzing protein families in the context of cancer development and progression. Specifically, our findings offer new insight into the potential of the serine/threonine kinase protein family as a target for therapeutic intervention in BC treatment. We identified five serine/threonine kinases as hub genes in BC and further found that functionally relevant domains, the binding and active sites, had high structural similarity. The serine/threonine kinase family includes other proteins and multiple binding and active sites, suggesting that our findings may generalize beyond the set of five examined here. The development of targeted therapies to specifically inhibit overexpression or activity of these proteins could lead to improved patient outcomes and a reduction in the toxicity often associated with traditional chemotherapy. The use of structural and sequence alignments in understanding the functionality of these families is a valuable avenue for exploration. Our findings suggest that *NEK2*, *MELK*, *PLK1*, *AURKB*, and *CHEK1* could be utilized for diagnostic screening, targeted inhibition therapy, and/or combination treatment monitoring while reducing side effects.

The study offers some insights into the function of serine/threonine kinases as central genes in BC; however, several limitations warrant attention. Firstly, the dependence on publicly accessible microarray and RNA-Seq datasets may introduce biases concerning sample selection, quality, or data processing, which could compromise the robustness and generalizability of the results. Secondly, although the study identifies gene regulatory

modules and hub genes, it primarily illustrates associations rather than establishing causality, necessitating additional experimental validation to confirm the causal roles of these genes in the progression of BC. Furthermore, technical variability between the microarray and RNA-Seq datasets may influence the outcomes, despite attempts to integrate them. The survival analysis linking elevated expression levels of hub genes to poorer prognoses does not provide direct clinical applicability, indicating a need for further clinical validation. Additionally, the classification of serine/threonine kinases as hub genes may oversimplify the intricate biology of BC, highlighting the necessity for more comprehensive investigations into the interactions of these kinases with other molecular pathways. Finally, while structural bioinformatics suggests potential drug targets, experimental validation is essential to confirm drug-target interactions, as computational models may not adequately reflect *in vivo* molecular complexities. These limitations underscore the importance of conducting further experimental and clinical studies to validate these findings and investigate the therapeutic potential of serine/threonine kinases in BC.

Conclusion

This comprehensive methodology facilitated the identification of five regulatory hub genes (*NEK2*, *MELK*, *PLK1*, *AURKB* (*STK12*), and *CHEK1*) that present significant promise as potential targets for pharmacological interventions. The hub genes identified belong to the serine/threonine kinase protein family, which encompasses enzymes responsible for the phosphorylation of serine or threonine residues in proteins, thus modulating numerous cellular functions. The identification of these hub genes is particularly significant, as they may serve as risk factors for BC, highlighting their essential role in the disease advancement. A notable discovery is that these hub genes exhibit structural similarities within their binding and catalytic domains. This structural resemblance suggests the possibility of developing therapeutic strategies capable of concurrently targeting all five genes. Such a strategy may prove to be more effective than focusing on individual genes, potentially resulting in enhanced treatment outcomes. Numerous serine/threonine kinase pathways have been associated with BC, including NEK kinases, PIM kinases, and the Akt/PKB signaling pathway. NEK kinases are implicated in the regulation of the cell cycle and mitotic processes, while PIM kinases are involved in cell survival and

proliferation. The Akt/PKB signaling pathway plays a vital role in cellular metabolism, growth, and survival. Targeting these pathways may provide new opportunities for BC treatment. By concentrating on these pivotal regulatory hubs, researchers and healthcare professionals may devise innovative treatment strategies that are more targeted and effective, ultimately leading to improved patient outcomes.

Authors' Contribution

M.SE: Conceptualization, Study design, bioinformatics analyses, statistical analysis, and drafting; M.GhZ: Conceptualization, Study design, bioinformatics analyses, statistical analysis, and drafting; Z.M: Bioinformatics analyses, data gathering, and reviewing the manuscript; M.Y: Bioinformatics analyses, data gathering, and reviewing the manuscript; A.B: Bioinformatics analyses, data gathering; E.NE: Data gathering, and reviewing the manuscript; H.Sh: Data gathering, and reviewing the manuscript; B.L: Data analysis, drafting and reviewing the manuscript. All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflict of Interest: None declared.

References

- 1 Coughlin SS. Epidemiology of Breast Cancer in Women. *Adv Exp Med Biol*. 2019;1152:9-29. doi: 10.1007/978-3-030-20301-6_2. PubMed PMID: 31456177.
- 2 Le Blanc JM, Heller DR, Friedrich A, Lannin DR, Park TS. Association of Medicaid Expansion Under the Affordable Care Act With Breast Cancer Stage at Diagnosis. *JAMA Surg*. 2020;155:752-8. doi: 10.1001/jamasurg.2020.1495. PubMed PMID: 32609338; PubMed Central PMCID: PMC7330827.
- 3 Brigham, Hospital Ws, 13 HMSCLP-PJKR, 25 GdaBCoMCCJDLA, Ilya IfSBR-SKRBBBBRETLJTVZWS. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490:61-70.
- 4 Koboldt D, Fulton R, McLellan M, Schmidt H, Kalicki-Veizer J, McMichael J, et al. MD Anderson Cancer Center.(2012). Comprehensive molecular portraits of human breast tumours. *Nature*.490:7418.
- 5 Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast

- cancer. *Nat Rev Dis Primers*. 2019;5:66. doi: 10.1038/s41572-019-0111-2. PubMed PMID: 31548545.
- 6 DeBerardinis RJ. Tumor Microenvironment, Metabolism, and Immunotherapy. *N Engl J Med*. 2020;382:869-71. doi: 10.1056/NEJMcibr1914890. PubMed PMID: 32101671.
 - 7 Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol*. 2007;25:1329-33. doi: 10.1200/JCO.2006.09.1066. PubMed PMID: 17416853; PubMed Central PMCID: PMCPMC2267287.
 - 8 Antoniou AC, Cunningham AP, Peto J, Evans DG, Lalloo F, Narod SA, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br J Cancer*. 2008;98:1457-66. doi: 10.1038/sj.bjc.6604305. PubMed PMID: 18349832; PubMed Central PMCID: PMCPMC2361716.
 - 9 Struwing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, et al. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N Engl J Med*. 1997;336:1401-8. doi: 10.1056/NEJM199705153362001. PubMed PMID: 9145676.
 - 10 Breast Cancer Association C, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Lucchini C, et al. Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. *N Engl J Med*. 2021;384:428-39. doi: 10.1056/NEJMoa1913948. PubMed PMID: 33471991; PubMed Central PMCID: PMCPMC7611105.
 - 11 Wilcox N, Dumont M, Gonzalez-Neira A, Carvalho S, Joly Beuparlant C, Crotti M, et al. Exome sequencing identifies breast cancer susceptibility genes and defines the contribution of coding variants to breast cancer risk. *Nat Genet*. 2023;55:1435-9. doi: 10.1038/s41588-023-01466-z. PubMed PMID: 37592023; PubMed Central PMCID: PMCPMC10484782.
 - 12 Shiovitz S, Korde LA. Genetics of breast cancer: a topic in evolution. *Ann Oncol*. 2015;26:1291-9. doi: 10.1093/annonc/mdv022. PubMed PMID: 25605744; PubMed Central PMCID: PMCPMC4478970.
 - 13 Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, et al. A Population-Based Study of Genes Previously Implicated in Breast Cancer. *N Engl J Med*. 2021;384:440-51. doi: 10.1056/NEJMoa2005936. PubMed PMID: 33471974; PubMed Central PMCID: PMCPMC8127622.
 - 14 Zhang D, Yang S, Li Y, Yao J, Ruan J, Zheng Y, et al. Prediction of Overall Survival Among Female Patients With Breast Cancer Using a Prognostic Signature Based on 8 DNA Repair-Related Genes. *JAMA Netw Open*. 2020;3:e2014622. doi: 10.1001/jamanetworkopen.2020.14622. PubMed PMID: 33017027; PubMed Central PMCID: PMCPMC7536586.
 - 15 Varn FS, Ung MH, Lou SK, Cheng C. Integrative analysis of survival-associated gene sets in breast cancer. *BMC Med Genomics*. 2015;8:11. doi: 10.1186/s12920-015-0086-0. PubMed PMID: 25881247; PubMed Central PMCID: PMCPMC4359519.
 - 16 Gyorffy B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Comput Struct Biotechnol J*. 2021;19:4101-9. doi: 10.1016/j.csbj.2021.07.014. PubMed PMID: 34527184; PubMed Central PMCID: PMCPMC8339292.
 - 17 Aubrey BJ, Strasser A, Kelly GL. Tumor-Suppressor Functions of the TP53 Pathway. *Cold Spring Harb Perspect Med*. 2016;6. doi: 10.1101/cshperspect.a026062. PubMed PMID: 27141080; PubMed Central PMCID: PMCPMC4852799.
 - 18 Deng CX. BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. *Nucleic Acids Res*. 2006;34:1416-26. doi: 10.1093/nar/gkl010. PubMed PMID: 16522651; PubMed Central PMCID: PMCPMC1390683.
 - 19 Bansal I, Pandey AK, Ruwali M. Small-molecule inhibitors of kinases in breast cancer therapy: recent advances, opportunities, and challenges. *Front Pharmacol*. 2023;14:1244597. doi: 10.3389/fphar.2023.1244597. PubMed PMID: 37711177; PubMed Central PMCID: PMCPMC10498465.
 - 20 Cioffi A, De Cobelli O, Veronesi P, La Vecchia C, Maisonneuve P, Corso G. Prevalence of Germline BRCA1/2 Variants in Ashkenazi and Non-Ashkenazi Prostate Cancer Populations: A Systematic Review and Meta-Analysis. *Cancers (Basel)*. 2023;15. doi: 10.3390/cancers15010306. PubMed PMID: 36612302; PubMed Central PMCID: PMCPMC9818251.
 - 21 Springer T. The Role of Reproductive Risk Factors in Hereditary Breast Cancer Within BRCA Mutation Carriers: Systematic Review and Meta-Analysis. Washington: Georgetown University; 2021.
 - 22 Moller NB, Boonen DS, Feldner ES, Hao Q, Larsen M, Laenkholm AV, et al. Validation of the BOADICEA model for predicting the likelihood of carrying pathogenic variants

- in eight breast and ovarian cancer susceptibility genes. *Sci Rep.* 2023;13:8536. doi: 10.1038/s41598-023-35755-8. PubMed PMID: 37237042; PubMed Central PMCID: PMCPMC10220031.
- 23 Xie Y, Shi H, Han B. Bioinformatic analysis of underlying mechanisms of Kawasaki disease via Weighted Gene Correlation Network Analysis (WGCNA) and the Least Absolute Shrinkage and Selection Operator method (LASSO) regression model. *BMC Pediatr.* 2023;23:90. doi: 10.1186/s12887-023-03896-4. PubMed PMID: 36829193; PubMed Central PMCID: PMCPMC9951419.
- 24 Asparuhova MB, Riedwyl D, Aizawa R, Raabe C, Couso-Queiruga E, Chappuis V. Local Concentrations of TGF-beta1 and IGF-1 Appear Determinant in Regulating Bone Regeneration in Human Postextraction Tooth Sockets. *Int J Mol Sci.* 2023;24. doi: 10.3390/ijms24098239. PubMed PMID: 37175951; PubMed Central PMCID: PMCPMC10179638.
- 25 Wei E, Hu M, Wu L, Pan X, Zhu Q, Liu H, et al. TGF-beta signaling regulates differentiation of MSCs in bone metabolism: disputes among viewpoints. *Stem Cell Res Ther.* 2024;15:156. doi: 10.1186/s13287-024-03761-w. PubMed PMID: 38816830; PubMed Central PMCID: PMCPMC11140988.
- 26 Sheynkman GM, Shortreed MR, Frey BL, Scalf M, Smith LM. Large-scale mass spectrometric detection of variant peptides resulting from nonsynonymous nucleotide differences. *J Proteome Res.* 2014;13:228-40. doi: 10.1021/pr4009207. PubMed PMID: 24175627; PubMed Central PMCID: PMCPMC3947302.
- 27 Qiu H, Li R, Li P, Xing W [Internet]. Expression data from esophageal squamous cell carcinoma patients. [cited 17 November 2020]. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE161533>
- 28 Chen YJ, Huang CS, Phan NN, Lu TP, Liu CY, Huang CJ, et al. Molecular subtyping of breast cancer intrinsic taxonomy with oligonucleotide microarray and NanoString nCounter. *Biosci Rep.* 2021;41. doi: 10.1042/BSR20211428. PubMed PMID: 34387660; PubMed Central PMCID: PMCPMC8385191.
- 29 Quigley DA, Tahiri A, Lüders T, Riis MH, Balmain A, Børresen-Dale A-L, et al. Abstract B59: Age and estrogen-dependent inflammation in breast adenocarcinoma and normal breast tissue. *Molecular Cancer Research.* 2016;14:B59-B.
- 30 Rajkumar T, Amritha S, Sridevi V, Gopal G, Sabitha K, Shirley S, et al. Identification and validation of plasma biomarkers for diagnosis of breast cancer in South Asian women. *Sci Rep.* 2022;12:100. doi: 10.1038/s41598-021-04176-w. PubMed PMID: 34997107; PubMed Central PMCID: PMCPMC8742108.
- 31 Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16:284-7. doi: 10.1089/omi.2011.0118. PubMed PMID: 22455463; PubMed Central PMCID: PMCPMC3339379.
- 32 Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559. doi: 10.1186/1471-2105-9-559. PubMed PMID: 19114008; PubMed Central PMCID: PMCPMC2631488.
- 33 Bawono P, Heringa J. PRALINE: a versatile multiple sequence alignment toolkit. *Methods Mol Biol.* 2014;1079:245-62. doi: 10.1007/978-1-62703-646-7_16. PubMed PMID: 24170407.
- 34 Krieger E, Vriend G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics.* 2014;30:2981-2. doi: 10.1093/bioinformatics/btu426. PubMed PMID: 24996895; PubMed Central PMCID: PMCPMC4184264.
- 35 Oshi M, Gandhi S, Huyser MR, Tokumaru Y, Yan L, Yamada A, et al. MELK expression in breast cancer is associated with infiltration of immune cell and pathological complete response (pCR) after neoadjuvant chemotherapy. *Am J Cancer Res.* 2021;11:4421-37. PubMed PMID: 34659896; PubMed Central PMCID: PMCPMC8493385.
- 36 Huang X, Zhang G, Tang T, Gao X, Liang T. One shoot, three birds: Targeting NEK2 orchestrates chemoradiotherapy, targeted therapy, and immunotherapy in cancer treatment. *Biochim Biophys Acta Rev Cancer.* 2022;1877:188696. doi: 10.1016/j.bbcan.2022.188696. PubMed PMID: 35157980.
- 37 Zhang X, Huang X, Xu J, Li E, Lao M, Tang T, et al. NEK2 inhibition triggers anti-pancreatic cancer immunity by targeting PD-L1. *Nat Commun.* 2021;12:4536. doi: 10.1038/s41467-021-24769-3. PubMed PMID: 34315872; PubMed Central PMCID: PMCPMC8316469.
- 38 Liu Z, Sun Q, Wang X. PLK1, A Potential Target for Cancer Therapy. *Transl Oncol.* 2017;10:22-32. doi: 10.1016/j.tranon.2016.10.003. PubMed PMID: 27888710; PubMed Central PMCID: PMCPMC5124362.
- 39 Equilibrina I, Krayukhina E, Inoue K, Ishikawa Y, Kawamoto A, Kato T, et al. The role of phosphorylation of histone H3 at serine 10

- in chromatin condensation in vitro. *Chromosome Science*. 2015;18:9-14.
- 40 Wu M, Pang JS, Sun Q, Huang Y, Hou JY, Chen G, et al. The clinical significance of CHEK1 in breast cancer: a high-throughput data analysis and immunohistochemical study. *Int J Clin Exp Pathol*. 2019;12:1-20. PubMed PMID: 31933717; PubMed Central PMCID: PMC6944032.
 - 41 Bhullar KS, Lagaron NO, McGowan EM, Parmar I, Jha A, Hubbard BP, et al. Kinase-targeted cancer therapies: progress, challenges and future directions. *Mol Cancer*. 2018;17:48. doi: 10.1186/s12943-018-0804-2. PubMed PMID: 29455673; PubMed Central PMCID: PMC5817855.
 - 42 Ghafouri-Fard S, Khoshbakht T, Hussien BM, Dong P, Gassler N, Taheri M, et al. A review on the role of cyclin dependent kinases in cancers. *Cancer Cell Int*. 2022;22:325. doi: 10.1186/s12935-022-02747-z. PubMed PMID: 36266723; PubMed Central PMCID: PMC9583502.
 - 43 Aquino-Acevedo AN, Orengo-Orengo JA, Cruz-Robles ME, Saavedra HI. Mitotic kinases are emerging therapeutic targets against metastatic breast cancer. *Cell Div*. 2024;19:21. doi: 10.1186/s13008-024-00125-x. PubMed PMID: 38886738; PubMed Central PMCID: PMC11184769.